1touch.io

Discovering Data Discovery in Depth

Data makes the world go around; enterprise applications, across cloud, web, mobile, and IoT are all data-dependent. But data is not a distinct entity floating in cyberspace. Data is a critical infrastructure that many of our processes, systems, and livelihoods depend on. Analyst IDC predicts that by 2025 this critical backbone of data will have increased tenfold from the 2016 figures to 163 zettabytes (a trillion gigabytes).

This data, created en masse by consumers and industry, across myriad devices, presents a challenge…How do you maintain privacy and security of data that is both massive and disparate?

Data Discovery-in-Depth provides the answer. In this short guide, we look at how data discovery, performed in an intelligent in-depth way, shines a spotlight on data to give you the visibility needed to make intelligent security and privacy decisions.

Why is 'Know Your Data' Crucial to Security and Privacy?

Knowing what data you are working with is essential when deciding upon the most appropriate security and privacy measures. Data, including Personally Identifiable Information (PII) has a complicated and expansive lifecycle. It touches many aspects of an enterprise.

Data privacy and security is not a tick box exercise; it is an operational and brand 'must-have'. Areas that are critically tied to the way an enterprise deals with its data include:

Trust and reputation

Data breaches breed distrust and negatively impact reputation. A Ponemon Institute study found 31% of consumers stopped using a company after a data breach.

Security

Data attracts cybercriminals like wasps to honey. In the first nine months of 2019, 7.9 billion data records were exposed, according to data from Risk Based Security.

Compliance and privacy

Data privacy is now heavily regulated. The EU data protection and privacy legislation, GDPR, is perhaps the most infamous data protection legislation. However, others, including the California Consumer Privacy Act (CCPA), are entering the compliance landscape, making data privacy a board-level issue. Fines for noncompliance are onerous. GDPR fines, to March 2020, reached almost half a billion euros.

However, to 'Know Your Data' you must first find it. Knowing the data you have, where it resides, and what part it plays in the overall enterprise data ecosystem is vital. Discovery-in-Depth provides the tools to truly Know Your Data.

What is Data Discovery-In-Depth?

The discipline of data discovery, when done well, informs how you locate, classify, and then protect data. Data discovery can be the all-seeing eyes of your data ecosystem. But in a disparate world of cloud, mobile, IoT, and edge apps, finding these data is not an easy exercise. The market entry of intelligent discovery techniques based on Artificial Intelligence has changed the way that data discovery works by expanding to provide an 'In-Depth' approach that is highly scalable.

Data Discovery-In-Depth tools provide the visibility needed to locate sensitive and personal data across the extended enterprise. But locating data is not enough. Data must also be understood. Modern **Data Discovery-In-Depth** also applies the context, relevance, and visualization needed to map security and privacy policies to data.

Typical techniques used to achieve a 'Discovery-In-Depth' approach include Artificial Intelligence (AI) and its subset, Machine Learning (ML), Natural Language Processing (NLP), and Network Analytics. It is this combination of powerful methodologies that adds 'in-depth' to data discovery.

Making the links across the enterprise to locate data is key; data hides and needs to be found by applying intelligent automation. Teasing out inherent and hidden relationships between data takes data discovery to new levels. By adding in context and relationships, organizations can ensure that their reach is deep and comprehensive.

How Does Discovery-in-Depth Help to Make Data Visible?

The discovery of data across extended IT networks should be seen as a process. It requires that several technologies come together, across a series of steps, to provide the thoroughness required for the task.

The process of **Data Discovery-in-Depth** is analogous to building a house. You start with a robust foundation, then add in the details, like windows and doors, as you build up the different parts of the house. When the house is built, you can truly see what you have and add the extra features like locks to doors, as needed.

1touch.io creates and maintains a full list of locations where an organization stores PII. These locations include on-premise (file servers, databases) and in the cloud (SaaS, AWS, Azure, etc). This situation is compounded by DevOps and the ubiquitous nature of ShadowIT; both see new PII locations PII in AWS, Virtual machines (VMs), etc appearing; these need to be accounted for. 1touch.io detects when new repositories come online, automatically adding them and spotting any potential PII in the repository traffic.

The key stages of data discovery-in-depth are:

Discover

This is your foundation stone. **Discovery-in-Depth** uses network and application analytics across your extended networks; this allows an accurate search for sensitive and personal information. An important first step in accurate and effective discovery is knowing where to look; 'drilling down' into an extended network to locate data and PII across the entire ecosystem of servers and cloud repositories is crucial; without this step, you will miss sensitive information.

Using identified PII candidates from data in motion allows visibilit of the storage location. Once these data locations are found, the **Discovery-in-Depth** work can begin in earnest.

In addition to locating all of your data, pattern recognition algorithms are also used to look for anomalies and duplicates of data. Visualization of these data is achievable using dashboards that give an at-a-glance view of the data ecosystem.

Analyze

Techniques such as Artificial Intelligence (AI), Machine Learning (ML), and NLP are used to analyze found data; this allows for the contextualization of the data.

Inventorize

The data can then be sorted. Duplicate data can be merged, and a data lineage presented graphically. Ultimately, this leads to the creation of master catalog of enterprise data.

Value

The master catalog is a valuable commodity. The catalog acts as a powerful tool for use in Know Your Data, data governance, and compliance; the data catalog gives you the intelligence needed to apply robust and effective security measures.
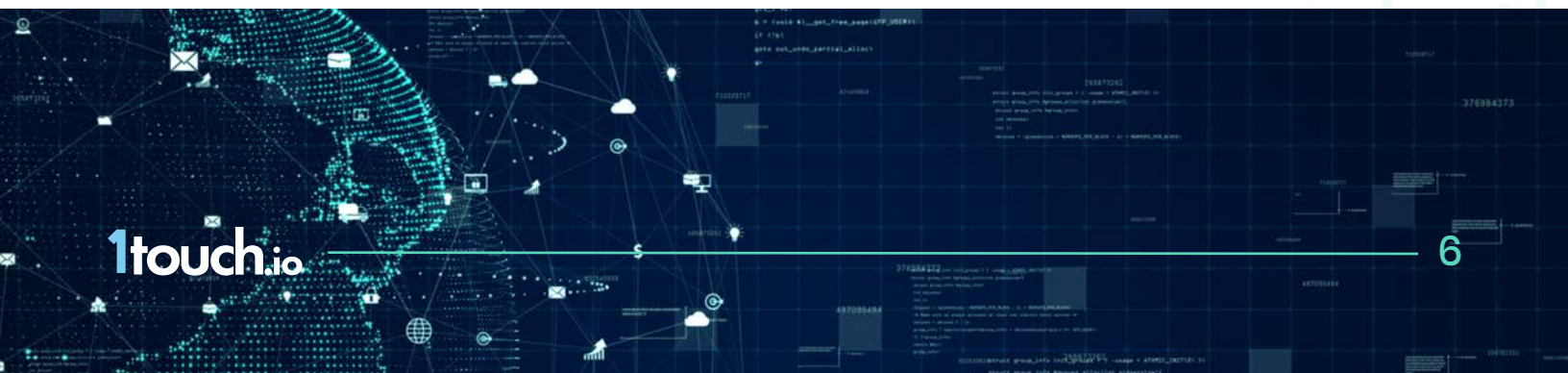
The Tentacles of Discovery-in-Depth

The importance of wide-area coverage in data discovery-in-depth cannot be overstated. All data must be discoverable. To do this requires a number of different approaches to locate data, including:

- **In motion and at rest:**
  Data exists in many states across its lifecycle. To discover data in motion, you need to make use of network sensors to locate all of your hidden and visible cloud repositories and servers.

- **Structured and unstructured:**
  During its lifecycle, data can be stored in different forms using varying techniques. Sensors that detect data in both SQL and NoSQL databases, as well as in a wide variety of document repositories, are vital to deep data discovery.

- **Known and unknown:**
  Data can go dark over time and be lost. A Data Discovery-in-Depth process must use intelligent techniques such as ML to locate even unknown data sources.

- **Within home-grown and off-the-shelf applications:**
  The average enterprise uses around 464 custom applications. A Discovery-in-Depth process must be flexible enough to reach out into home-grown applications as well as off-the-shelf solutions.

How Discovery-In-Depth Helps in Compliance

Meeting the requirements of regulations such as CCPA, GDPR, HIPAA, etc., that deal with data security and privacy, can be both nuanced and intensive in preparation and implementation. However, the results of a Data Discovery-in-Depth process can be invaluable in meeting that compliance. Knowing where your enterprise data lies, as well as its context, is the starting point in meeting the requirements of data protection laws.

Conclusion

Enterprise data is fluid. It moves across many apps, through multiple cloud infrastructures, and can often feel outside of our control. However, the use of **Data Discovery-in-Depth**, which is comprehensive enough to find all PII, no matter where it is located, gives us the means to govern these data. And, the application of an intelligent automated system to find and classify data within a dedicated process of discovery is an important tool to use in compliance.

Let **Discovery-In-Depth** be your eyes to make your enterprise data visible, no matter what or where it is.